Supplemental Information: Aerobic fitness and mnemonic discrimination across the adult lifespan

Material and Methods

Cardiorespiratory Fitness Test Procedure

Participants completed a submaximal-graded exercise test on a motor-driven treadmill (Precor TRM835) using a modified Balke protocol to estimate CRF (ACSM, 2014; Cooper & Storer, 2001; Hagberg, 1994). We determined CRF by estimating maximal oxygen uptake ($\dot{V}O_{2MAX}$) based on the linear relationship between heart rate and oxygen uptake ($\dot{V}O_2$) (Wasserman et al., 2012). At each minute of the test, oxygen uptake was predicted from treadmill speed and grade using standard, published equations (ACSM, 2014). After a 3-minute warm-up, participants walked at a constant speed as the grade of the treadmill increased incrementally. We continuously recorded heart rate in beats per minute as participants walked, and the test concluded when the subject reached 85% of age-predicted maximal heart rate (HR_{max}; Tanaka et al., 2001). We conservatively selected this termination criterion as a safety precaution for older adults. Participants were asked to rate their perceived level of exertion every three minutes using the Borg 6-20 scale (Borg, 1982). Using the American College of Sports Medicine metabolic equation for gross $\dot{V}O_2$ for walking (Equation 1), we calculated oxygen uptake for each minute of the test as follows:

$$\dot{V}O_2 = 0.1\text{mL/kg/min} \times S + 1.8\ \text{mL/kg/min} \times S \times G + 3.5\ \text{mL/kg/min} \quad \text{(Equation 1)}$$

$\dot{V}O_2$, S, G, and 3.5 represent gross oxygen uptake (ml/(kg ·min)), treadmill speed (m/min), the grade percent of the treadmill expressed as a fraction, and the resting oxygen uptake, respectively (ACSM, 2014). The calculated $\dot{V}O_2$ was then plotted with its corresponding HR at the end of each minute. Subsequently, we calculated a linear regression using individual subject's predicted HR_{max} (Tanaka et al., 2001), allowing us to estimate maximal aerobic capacity as our measure of CRF. This estimated $\dot{V}O_{2MAX}$ serves as the measure of CRF level throughout the statistical analyses. Following CRF testing, participants were given a three to five minute cool-down, and HR and blood pressure were monitored until they returned to resting levels. Eighteen participants did not reach termination criteria (85% HR_{MAX}, leaving sixty-two participants with CRF data for inclusion in the statistical analyses. We note the limitation of the use of a submaximal exercise

test. Submaximal CRF testing relies on estimation of $\dot{V}O_{2MAX}$ as well as the two factors used for that estimation ($\dot{V}O_2$ for speed and grade, and predicted HR_{MAX} based on age), thus likely introducing additional variability in our CRF assessment. However, submaximal exercise tests are a reliable and established way to estimate CRF without the need for high-intensity testing that may be unsuitable and unsafe for older adult populations, such as those in our study (ACSM, 2014).

Detailed Neuropsychological Assessment Results

We found no differences between men and women within each age group for TMT ratio (YA_{women} = 2.18 ± .70, YA_{men} = 2.17 ± .50, t(12.5) = .03; MY_{women} = 2.4 ± .60, MY_{men} = 2.09 ± .65, t(20.2) = 1.20; MO_{women} = 2.22 ± .74, MO_{men} = 2.22 ± .56, t(18.4) = .01; OA_{women} = 2.56 ± .84, OA_{men} = 2.22 ± .71, t(14.6) = .92; all *p*s > .05), VST ratio (YA_{women} =1.69 ± .32, YA_{men} = 1.54 ± .22, t(12.23) = 1.12; MY_{women} = 1.74 ± .42, MY_{men} = 1.56 ± .34, t(17.0) = 1.11; MO_{women} = 1.95 ± .37, MO_{men} = 1.93 ± .43, t(18.0) = .09; OA_{women} = 2.13 ± .52, OA_{men} = 2.07 ± .61, t(11.0) = .83; all *p*s > .05), or the SBSOD scale (YA_{women} =4.90 ± .83, YA_{men} = 4.84 ± .98, t(15.0) = .13; MY_{women} = 5.04 ± 1.3, MY_{men} = 4.65 ± 1.2, t(18.7) = .75; MO_{women} = 4.91 ± .85, MO_{men} = 5.47 ± .82, t(18.9) = -1.5; OA_{women} = 4.87 ± .73, OA_{men} = 5.19 ± .80, t(11.7) = -.87; all *p*s > .05). For RAVLT, we found no sex differences in the two youngest age groups (RAVLT Total: YA_{women} = 60.6 ± 5.55, YA_{men} = 56.7 ± 9.0, t(13.5) = 1.1; MY_{women} = 59.9 ± 9.10, MY_{men} = 56.7 ± 10.5, t(20.7) = .78l; RAVLT Immediate: YA_{women} = 13.0 ± 2.2, YA_{men} = 12.1 ± 3.6, t(13.5) = .63; MY_{women} = 12.7 ± 2.11, MY_{men} = 12.6 ± 3.07, t(20.8) = .08; RAVLT Delay: YA_{women} = 13.1 ± 2.17, YA_{men} = 12.7 ± 2.83, t(14.7) = .38; MY_{women} = 13.0 ± 1.89, MY_{men} = 12.0 ± 3.16, t(20.0) = .94). Furthermore, men in the MO age group performed significantly worse on all indices of RAVLT (RAVLT Total: MO_{women} = 57.5 ± 6.04, MO_{men} = 47.1 ± 7.28, t(17.6) = 3.56, $p < .01$; RAVLT Immediate: MO_{women} = 12.3 ± 2.33, MO_{men} = 8.3 ± 3.4, t(15.7) = 3.09, $p < .01$; RAVLT Delay: MO_{women} = 12.5 ± 1.81, MO_{men} = 8.5 ± 4.17, t(12.0) = 2.84, $p < .05$; see Table 1 in main text). In our oldest age group, men did significantly poorer on RAVLT total (OA_{women} = 50.8 ± 7.4, OA_{men} = 39.6 ± 5.2, t(16.2) = 3.9, $p < .01$), but the sex difference between men and

women in this age group only reached marginal significance for the RAVLT Immediate (OA_{women} = 10.8 ± 2.34, OA_{men} = 8.0 ± 2.89, t(10.6) = 1.75, p = .11) or RAVLT Delay (OA_{women} = 9.33 ± 3.11, OA_{men} = 6.57 ± 3.05, t(12.9) = 1.89, p = .08) separately. Participant performance on these neuropsychological measures fell within the range for normative data (Strauss et al., 2006). We present these results in Table 1 for a thorough characterization of our study's sample.

Table S1. Response proportion for MST by stimulus condition

		Stimulus Condition		
Age Group	Response	Target	Lure	Foil
Young Adult (YA)	Old	.81 (.12)	.36 (.07)	.06 (.07)
	Similar	.13 (.08)	.53 (.13)	.12 (.07)
	New	.06 (.09)	.11 (.09)	.82 (.11)
Middle-Young (MY)	Old	.84 (.12)	.45 (.16)	.04 (.04)
	Similar	.10 (.08)	.43 (.19)	.15 (.10)
	New	.06 (.07)	.12 (.08)	.81 (.13)
Middle-Older (MO)	Old	.88 (.09)	.54 (.18)	.05 (.05)
	Similar	.08 (.06)	.36 (.20)	.21 (.19)
	New	.04 (.07)	.10 (.06)	.74 (.19)
Older Adult (OA)	Old	.82 (.12)	.56 (.12)	.08 (.11)
	Similar	.10 (.09)	.23 (.15)	.11 (.11)
	New	.08 (.07)	.21 (.12)	.81 (.16)

Average percent (sd) endorsed for the MST for each age group by stimulus condition and response.

Table S2. Average performance on the MST by age group and CRF level

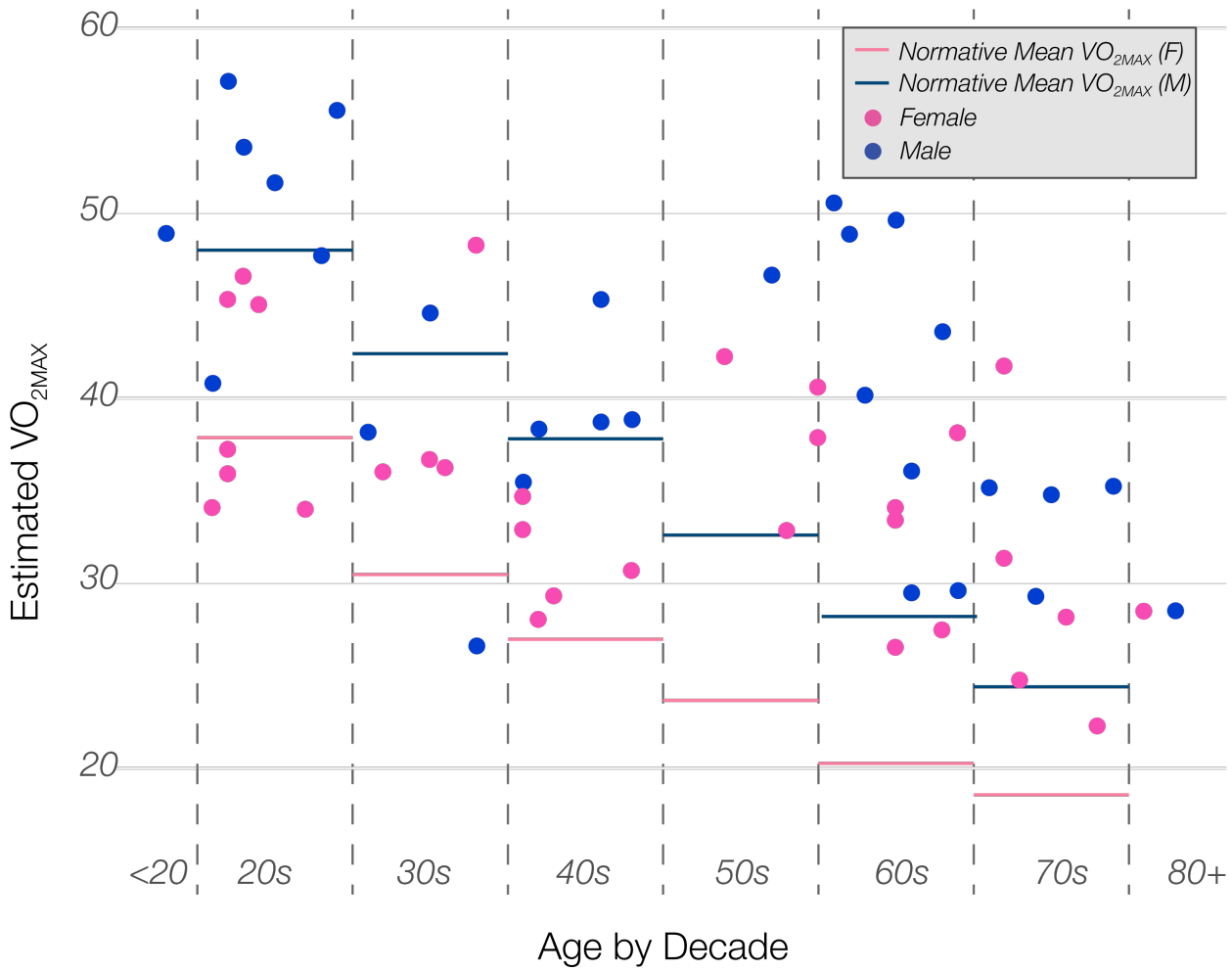
	Young Adult (YA) Ages 18-30	Middle-Young (MY) Ages 31-49	Middle-Older (MO) Ages 50-66	Older Adult (OA) Ages 67 - 85
All Subjects				
LDI	.41 (.16)	.28 (.22)	.14 (.26)	.06 (.25)
REC	.76 (.15)	.80 (.14)	.83 (.09)	.74 (.17)
Higher-fit				
LDI	.41 (.12)	.21 (.19)	.12 (.37)	.11 (.20)
REC	.79 (.08)	.85 (.08)	.85 (.06)	.81 (.09)
Lower-fit				
LDI	.37 (.19)	.31 (.15)	.13 (.25)	.03 (.31)
REC	.73 (.20)	.79 (.14)	.84 (.08)	.76 (.16)

Average scores (sd) for lure discrimination index (LDI) and recognition memory (REC) for each age group and subsequently separated out by CRF level (higher-fit vs. lower-fit).

Table S3. Mean accuracy and response times for the route disambiguation task.

	Non-Overlapping			Overlapping		
	CP1	*CP2*	*CP3*	*CP1*	*CP2*	*CP3*
Accuracy						
Young Adults (YA)	.75 (.17)	.79 (.16)	.84 (.13)	.80 (.14)	.67 (.08)	.64 (.20)
Middle-Young (MY)	.72 (.20)	.72 (.23)	.76 (.24)	.76 (.20)	.61 (.21)	.63 (.21)
Middle-Older (MO)	.59 (.25)	.65 (.23)	.69 (.21)	.67 (.19)	.48 (.15)	.56 (.17)
Older Adults (OA)	.50 (.14)	.50 (.16)	.54 (.15)	.53 (.22)	.36 (.08)	.46 (.09)
Response Times						
Young Adults (YA)	987.4 (262.9)	499.2 (193.4)	434.8 (182.7)	1006.4 (292.3)	631.9 (263.1)	524.2 (227.2)
Middle-Young (MY)	1087.4 (466.8)	642.8 (342.0)	543.2 (262.9)	1171.3 (428.1)	728.2 (317.7)	599.2 (254.9)
Middle-Older (MO)	1351.1 (391.0)	803.1 (270.4)	706.0 (216.6)	1403.5 (403.5)	954.0 (371.6)	733.9 (249.5)
Older Adults (OA)	1793.5 (552.8)	1199.5 (388.8)	969.6 (233.3)	1777.2 (567.6)	1247.1 (450.6)	1008.7 (262.2)

Means (sd) for accuracy and response times across non-overlapping and overlapping choice points (CPs) by age group for the route disambiguation task. Overlapping choice points 2 and 3 require the greatest spatial disambiguation.



Supplemental Figure 1. Estimated $\dot{V}O_{2MAX}$ by age and sex compared to normative mean. Each individual point represents estimated $\dot{V}O_{2MAX}$ for an individual. Pink colors reflect women, blue reflect men. Pink and blue horizontal lines in each decade denote the normative mean by sex. Data from younger and middle-aged adults distribute well around the age and sex-adjusted normative mean. Older adults in our sample tended to be higher fit than the age and sex-adjusted normative mean.